

NUMEROSITY DISCRIMINATION IN PRESCHOOL CHILDREN

ALZIRA ALMEIDA, JOANA ARANTES, AND ARMANDO MACHADO

UNIVERSITY OF MINHO, PORTUGAL

We used a numerical bisection procedure to examine preschool children's sensitivity to the numerical attributes of stimuli. In Experiment 1 children performed two tasks. In the Cups Task they earned coins for choosing a green cup after two drumbeats and a blue cup after eight drumbeats. In the Gloves Task they earned coins for raising a red glove on their left hand after two drumbeats and a yellow glove on their right hand after eight drumbeats. Then in each task a psychometric function was obtained by presenting intermediate numerosities and recording the percentage of trials in which children chose the "many" option. In Experiment 2 children's performance in a "2 vs. 8" discrimination was compared with their performance in a "4 vs. 16" discrimination. Results showed that the individual psychometric functions were of two types, one in which the percentage of "many" choices increased gradually with stimulus numerosity and another in which it increased abruptly, in a step-like manner. Although the average point of subjective equality was close to the geometric mean of the anchor numerosities and the average functions for "2 vs. 8" and "4 vs. 16" superimposed when plotted on a common scale (the scalar property), the individual data were highly variable both across tasks (Cups and Gloves) and numerosity ranges ('2 vs. 8' and '4 vs. 16'). It is suggested that between- and within-subjects variability in the psychometric function is related to children's verbalizations about the sample stimulus.

Key words: bisection procedure, numerosity discrimination, point of subjective equality, psychometric function, scalar property, children

Nonhuman animals and humans display different levels of numerical competence. One of the most primitive of these levels is numerosity discrimination, the ability to distinguish many from few (Davis & Perusse, 1988; Emmerton, 2001). A set with many elements may be the discriminative stimulus for a rat's response of pressing the left lever and a set with few elements the discriminative stimulus for a rat's response of pressing the right lever. The elements of the set may be stimuli (e.g., Meck & Church, 1983) or responses (e.g., Fetterman, 1993; Rilling 1967; Rilling & McDiarmid, 1965). In a procedural variant, the sample set comprises two types of stimuli such as two lights of different colors, each with a different numerosity. After the sample, the animal must choose one of two responses according to the relative numerosities of the two colors (Honig & Matheson, 1995; Honig & Steward, 1989; Keen & Machado, 1999; Machado & Keen, 2002). A

large number of studies show that this level of numerical competence is present in mammals and birds (Gallistel, 1990; Rilling, 1993; Roberts, 1998; Shettleworth, 1998).

A more sophisticated level of numerical competence involves the operational concept of number studied by Piaget and other researchers (e.g., Piaget, 1952; see also Gallistel & Gelman, 1992; Gelman & Gallistel, 1978). That level is defined by the understanding of the logical ideas of class inclusion (a set with one element is included in a set with two elements, which in turn is included in a set with three elements, etc.), seriation of quantity (1 is less than 2, which is less than 3, etc.), and number conservation (the number of items in a set is independent of the spatial disposition of the items). If there is little doubt that this level of numerical competence is achieved by most humans typically around 6 to 8 years of age, there is doubt whether any nonhuman animal can reach it (for comparisons of numerical competence in animals, infants, and children see Brannon & Roitman, 2003; Emmerton, 2001; Gallistel & Gellman, 1992; Mix, Huttenlocher, & Levine, 2002).

It is commonly believed that the different levels of numerical competence are interrelated, although this idea is rarely articulated in theory. To illustrate, consider the idea of serial order. The abilities involved in ordering sets

The work reported in this study was part of the Master's thesis defended by the first author at the University of Minho, Portugal. Research was supported by a grant from the Portuguese Foundation for Science and Technology (FCT) to Armando Machado. Address correspondence to (E-mail: armandom@iep.uminho.pt). We thank Francisco Silva and Luis Oliveira for their comments on earlier versions of the paper.

doi: 10.1901/jeab.2007.88-339

according to number and drawing inferences from the resulting order (e.g., if $A < B$ and $B < C$, then $A < C$) may ultimately rest on the primitive ability to discriminate the numerosity of two sets. Conversely, it is hard to imagine the mastery of the number concept, or the ability to count, without the primitive ability to discriminate numerosity (Mix *et al.*, 2002). Hence, the most general goal of the present study was to characterize numerosity discrimination in children. A good grasp of one of the foundational stones of the concept of number is necessary to understand the development and learning of numerical competence.

The study of numerosity discrimination in animals has progressed considerably and for that reason the procedures used and the results obtained with them may be good starting points for the corresponding studies with young children. Consider the bisection procedure mentioned above. The animal is exposed to one of two sample numerosities, say, a sequence of $N_F = 2$ tones or a sequence of $N_M = 8$ tones (subscripts “F” and “M” stand for samples with “few” and “many” stimuli, respectively). At the end of the sample, the animal must choose between two comparison stimuli in order to get food. A pigeon, for example, may receive food for choosing a red key following two tones and a green key following eight tones. In this case, the red and green stimuli may be referred to as the “few” and “many” alternatives, respectively. After the animal learns the basic discrimination, the experimenter introduces generalization tests, that is, samples with intermediate numerosities ($2 < N < 8$) and records the percentage of choices of the “many” key following each numerosity. Plotting the percentage of “many” choices against stimulus numerosity yields a psychometric function.

A typical psychometric function has three properties (e.g., Brannon & Roitman, 2003; Emmerton, 2001; Meck & Church, 1983; Roberts, 1998; Shettleworth, 1998). First, it is a monotonic increasing function of stimulus numerosity, a function that starts close to 0 percent for the smallest numerosity (N_F) and ends close to 100 percent for the largest numerosity (N_M). Most psychometric functions are S-shaped, with a positively accelerated initial segment and a negatively accelerated final segment. Second, the numerosity at which the psychometric function equals 50

percent defines the bisection point or the Point of Subjective Equality, abbreviated PSE. In most studies with animals, the PSE is close to the geometric mean of the two training samples, that is, $PSE = \sqrt{N_F \times N_M}$. In the example above with $N_F = 2$ and $N_M = 8$, the PSE would be located at 4. And third, the psychometric functions obtained with different pairs of sample numerosities, but with the ratio between the training numerosities held constant, superimpose when plotted on a common axis. To illustrate, suppose that the same pigeon that learned to discriminate two tones from eight tones, subsequently learns to discriminate 4 tones from 16 tones. To plot on the same axis the two psychometric functions (one obtained with numerosities ranging from 2 to 8 and the other obtained with numerosities ranging from 4 to 16), one would rescale the numerosities from the ‘4 vs. 16’ set by dividing them by 2. The two functions would then superimpose. Superimposition suggests a property similar to Weber’s law for numerosity discrimination in the sense that equal numerosity ratios yield equal discriminabilities.

The question naturally arises as to whether numerosity discrimination in children shares the same properties as numerosity discrimination in animals. However, only two studies have applied the numerosity bisection procedure to children and studied the properties of the psychometric function obtained with that procedure. In one of them (Droit-Volet, Clément, & Fayol, 2003), children (5- and 8-year olds) and adults were studied in a bisection task with a sequence of stimuli as the sample. During training, the number of stimuli comprising the sample was perfectly correlated with the duration of the sample (e.g., two circles flashed in a computer monitor for 2 s or eight circles flashed for 8 s). The participants were instructed to pay attention to the number of stimuli and ignore their duration. On test trials, the authors dissociated duration from number by presenting samples with variable numbers of stimuli (from 2 to 8) and constant duration (4 s), or samples with variable duration (from 2 s to 8 s) and constant numbers of stimuli (4). The goal was to determine whether children and adults differ in resistance to interference by the irrelevant stimulus dimension (time in this case).

Results showed that duration did not interfere with the numerical discrimination for any age group. When the sample duration varied while its numerosity remained constant, the group psychometric functions remained roughly horizontal, but when sample numerosity varied while its duration remained constant, the group psychometric functions had the usual S-shaped, monotonic increasing trend. The average PSEs for three age groups were closer to the arithmetic mean of the anchor numerosities (5) than to their geometric mean (4).

The other study (Jordan & Brannon, 2006) used a numerosity discrimination task with 6-year-old children and compared the results with those produced by rhesus monkeys on the same task. Specifically, children and monkeys saw one of two stimulus samples on a touch sensitive screen (e.g., two or eight circles within a rectangle). After they touched the screen, the sample disappeared and two comparison rectangles were shown, one with two and the other with eight circles. The subjects received a reinforcer for choosing the comparison that matched the sample. On probe trials, the sample numerosity varied from 3 to 7, but the subject still had to choose which comparison rectangle (2 or 8) was more similar to the sample. To study the superimposition property, the authors used two pairs of training samples, '2 vs. 8', as described above, and '3 vs. 12'.

Results showed that the group psychometric functions increased initially almost linearly and then in a negatively accelerated way. The PSEs, obtained by fitting a cumulative Gaussian distribution to the group functions and retrieving its estimated mean, equaled 3.53 ± 0.15 in the '2 vs. 8' and 4.96 ± 0.20 in the '3 vs. 12' condition, in both cases values close to, but significantly below, the geometric means of 4 and 6. The two group psychometric functions overlapped when all sample numerosities were expressed as proportions of the PSEs. Finally, the data from children and rhesus monkeys also overlapped.

The two studies left several questions unanswered. First, because both studies reported only average psychometric functions, we do not know the extent to which group data represent individual data or how well the PSE estimated from fitting the average curve represents the average of the PSEs obtained

from fitting the individual curves. The issue of group versus individual data (see Sidman, 1960) is particularly relevant with regard to the superimposition property because it is possible to have superimposition at the group level without any superimposition at the individual level. Second, even at the group level, it remains unclear why Jordan and Brannon (2006) obtained a PSE close to the geometric mean whereas Droit-Volet et al. (2003) obtained a PSE close to the arithmetic mean. One hypothesis is that the geometric mean may be the modal result when the items that comprise the sample are presented simultaneously (e.g., two or eight circles displayed on a touch screen all at the same time, as Jordan and Brannon did) and the arithmetic mean the modal result when the items that comprise the sample are presented successively (e.g., two or eight circles flashed on a screen one at a time, as Droit-Volet et al. did). We refer to the two tasks as *simultaneous* and *successive* numerosity discrimination tasks, respectively. Third, superimposition has been shown with children only in simultaneous tasks, that is, in tasks in which the sample elements are presented all at once (Jordan & Brannon). Therefore, whether it holds in successive tasks in which the sample elements are presented serially, one at a time, remains to be seen. Fourth, and finally, it is also unclear whether the results obtained by Droit-Volet et al. in a successive task, particularly with the 5-year-old children, would be obtained if the experimenters did not instruct them to attend to number and ignore duration. In fact, the study had one additional feature that is important in the present context: Half of the participants were instructed to count during the sample and the other half were explicitly instructed not to count and instead to repeat aloud, as fast as possible, "blabla". Whether the results hold when the children's behavior during the sample is not manipulated also remains to be determined.

Two experiments reported below attempted to clarify the foregoing issues. In Experiment 1, children learned to discriminate between two and eight drumbeats in two different tasks, and then they were exposed to intermediate numerosities to obtain a psychometric function for each task. Some of the children learned subsequently in Experiment 2 to discriminate 4 from 16 drumbeats and then also produced a psychometric function. By comparing group

and individual psychometric functions we attempted to determine whether the group functions represent well the individual functions and whether the average PSE is close to the geometric mean (as the animal studies suggest and Jordan & Brannon, 2006, obtained) or the arithmetic mean (as Driot-Volet *et al.*, 2003, obtained). By comparing the functions obtained with the numerosity pairs '2 vs. 8' and '4 vs. 16' we addressed the issue of superimposition in successive tasks. And because the experimenter made no reference to sample attributes (number or duration) or to counting/not counting behavior during the sample, we attempted to determine whether such instructions change performance in successive numerosity discrimination tasks.

EXPERIMENT 1

Two tasks were designed to study numerosity discrimination in preschool children. In the Cups Task children learned to choose a green cup following one sample numerosity and a blue cup following another sample numerosity. In the Gloves Task they learned to raise a red glove on the left hand following one sample numerosity and a yellow glove on the right hand following another sample numerosity. They received one coin after each correct response and lost one coin after each incorrect response. After the children learned the basic discrimination they were presented with intermediate numerosities in order to obtain a psychometric function.

The use of two tasks allowed us to explore which one would be easier for preschool children. In the Cups Task, the correct response was signaled exclusively by the color of the cup; the cups were equal in all other respects and their left-right position varied randomly across trials. In the Gloves Task, color and position were correlated and therefore children could base their choices on either dimension. Hence, the Gloves Task was presumably easier than the Cups Task. The use of two tasks enables us also to test the consistency of the results across tasks.

METHOD

Subjects

The sample consisted of 19 children, 8 girls and 11 boys. The average age was 5 years and

9 months (range: from 5 years and 5 months to 6 years and 3 months). The children came from two preschools located in the city of Braga, Portugal. Parents and the school principal gave informed consent for the children's participation in the study.

Materials

The following objects were used during the experiments: (a) one small drum, 20 cm in diameter and 15 cm high, with two 15-cm drumsticks attached to it; the drum was used to generate sounds; (b) two opaque drinking cups, one green and the other blue, 9 cm high; the cups served as comparison stimuli; (c) two wool gloves, one red and the other yellow, of adequate size to a 5- to 6-year-old child; the gloves also were used as comparison stimuli; and (d) a bag with plastic coins that served as rewards. At the end of each session the coins could be exchanged for chocolate bars.

Procedure

Children participated in the Cups Task and the Gloves Task in a counterbalanced order. In both cases, the children came to a small room in the school and sat either at a round table or on the floor facing the experimenter. After a few minutes of informal interaction with the child, the experimenter invited the child to play a game.

Cups Task. The instructions (translated from the Portuguese) were as follows:

"I am going to teach you a very funny game. If you want to play it well, you have to pay attention to what I say. Let's pretend you're a little boy (girl) living on an island. You have to eat a lot to fight against mean animals that live around you. But you eat only chocolates... only chocolates give you energy and strength to defeat the mean animals. On the island lives a friendly bear that speaks through a drum. He is going to help you find chocolates. Let's pretend I am the bear, OK? Now, pay attention to see how you can find chocolates. Do you see these cups? Which color is this one? Very good. And this one? Very good. I am going to turn them upside down. Underneath one of them I am going to hide one coin [the experimenter shows a coin taken from a plastic bag placed on an adjacent table]. You need to close your eyes and not open them until I say

so, ok? No peeking, please. [The child closes his eyes and covers them with his hands while the experimenter hides the coin]. Now, pay attention to what I am going to tell you with my drum. When I say this [the experimenter hits the drum twice, about one drumbeat per second] you raise the cup that you think I told you, and when I say this [the experimenter hits the drum 8 times, again about one drumbeat per second] you raise the other cup. You need to pay attention to what I say with the drum because it is the only way you have to know where the coin is. Did you understand? If you raise the cup with the coin you win the coin and keep it, but if you raise the wrong cup then you need to give me one of your coins [provided the child had at least one coin]. The more coins you have at the end of the game, the more chocolates you can win [the chocolate bars were not visible during the game]. Shall we play the game?"

During training, the child learned to associate one of the cups with two drumbeats (the "few" numerosity) and the other cup with eight drumbeats (the "many" numerosity). The cup assignment was counterbalanced across children, but for clarity we describe the procedure and the results as if all children had the assignment 'green cup after two drumbeats and blue cup after eight drumbeats'. The two cups were presented the same number of times on the left and right positions, about 20 cm aside. The training lasted until children had completed a minimum of 20 trials and made five consecutive correct choices following each stimulus numerosity ($p < .05$ for each sample).

If after 20 trials a child had not learned to discriminate the two numerosities, the experimenter simplified the task by introducing easier trials. For example, after an incorrect choice, the trial was repeated (correction trials method) and, if errors persisted, the left-right location of the cups was not changed. Once the child learned the discrimination with the simpler trials, the experimenter returned to the original ones.

After the basic discrimination was learned, the test phase began. It was composed of three series of 30 trials each. Each series included in pseudorandom order the two anchor numerosities of 2 and 8, each presented five times, and the intermediate numerosities of 3, 4, 5, 6 and 7, each presented four times. Following

the anchor numerosities, only the choice of the correct cup was rewarded, but following the intermediate numerosities any choice was rewarded (i.e., both cups hid one coin).

Gloves Task. The experimenter asked half of the children to put the red glove on the left hand and the yellow glove on the right hand, and the other half to put on the gloves in the opposite order. The instructions, training, and testing phases were identical to the *Cups Task* except that instead of choosing one of the cups after hearing the drumbeats (also with their eyes closed), the children were asked to raise one hand. If the choice was correct, the child won one coin, and if it was incorrect, the child lost one coin (provided she had at least one coin). The assignment of the correct gloves to the two sample numerosities was counterbalanced across children, but for clarity we describe the procedure and results as if all children learned to raise the red glove after two drumbeats and the yellow glove after eight drumbeats.

At the end of the first task that each child performed, the experimenter asked a few questions to determine if the child knew how to count: "How many coins did you win?", "Please give me three coins", "Can you make a pile with five coins?", "And one with eight coins?"

Data analysis. Usually the psychometric functions obtained with animals, either in studies of timing or numerosity discrimination, are well described by a cumulative normal distribution (e.g., Meck & Church, 1983; Gibbon, 1981) or by a logistic curve (e.g., Keen & Machado, 1999; Machado & Keen, 2002). We chose the logistic because it is mathematically more tractable. Its equation is, $P(\text{"many"} | n) = 100 / (1 + \exp(-\lambda(n - \mu)))$, where $P(\text{"many"} | n)$ is the percentage of "many" choices following a sample of numerosity n , $\mu > 0$ is the PSE (hence $P(\text{"many"} | \mu) = 50\%$), and $\lambda > 0$ is related to the slope of the function at the PSE. For relatively small values of λ , the function increases gradually from about 0 to about 100, whereas for relatively large values of λ , the function increases abruptly, in a step-like manner, with the step centered at stimulus numerosity μ . The value of λ measures the subject's sensitivity to numerosity, with higher values meaning greater sensitivity. Parameter λ is inversely proportional to another common measure of

sensitivity, the Weber ratio (i.e., Weber ratio $\approx 0.55/\lambda$). Parameters μ and λ are also referred to as the location and scale parameters, respectively, because changes in μ displace the curve horizontally and changes in λ change the scale of the independent variable.

RESULTS AND DISCUSSION

Fifteen of the 19 participants completed the Gloves Task. Of these, 13 learned the task during the first 20 trials and 2 needed seven or nine additional trials to meet the learning criterion. In the Cups Task, of the 18 children that completed it, 13 learned the task during the first 20 trials and the remaining 5 needed from 3 to 30 additional trials to meet the learning criterion. Children who had some difficulties with the discrimination based their first choices on the left–right position of the cups or alternated their choice on successive trials (green cup, blue cup, green cup, and so on). However, at the end of the training phase all children discriminated perfectly two from eight drumbeats in both tasks. Contrary to expectation, there was no evidence that the Gloves Task was easier to learn than the Cups Task.

All remaining analyses focus on the generalization data. We start with the average data for the two tasks to compare them with the results reported in other studies; we then analyze the individual data. Three main issues are considered: the general form of the psychometric function; its slope and location (PSE) parameters; and the consistency of the individual data across tasks.

Figure 1 shows the group psychometric functions for the Gloves Task and Cups Task. In both, the percentage of “many” choices increased monotonically and in a negatively accelerated way with the number of drumbeats. The function for the Gloves Task had a steeper slope and a smaller PSE than the function for the Cups Task (λ : 1.3 vs. 1.1; μ : 3.6 vs. 4.1), but the differences between the two functions were small.

For comparison purposes, Figure 1 also shows the data from Droit-Volet *et al.* (2003) for the 5-year-old, noncounting group, and from Jordan and Brannon (2006). The function for the Gloves Task was close to the function obtained by Jordan and Brannon. The authors report a PSE equal to 3.53 and a Weber ratio equal to 0.24; the corresponding

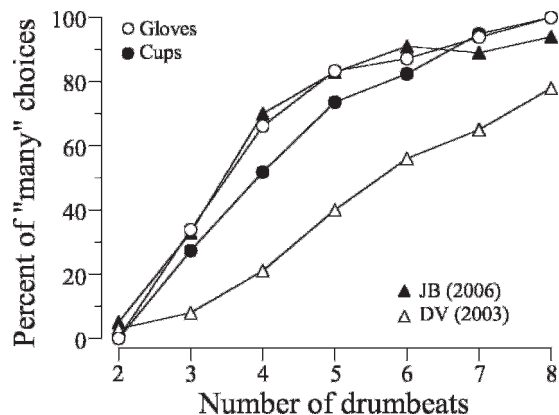


Fig. 1. Average psychometric functions in the Cups Task and Gloves Task of Experiment 1 and in two other studies, Droit-Volet, Clément, & Fayol (2003), abbreviated DV (2003), and Jordan & Brannon (2006), abbreviated JB (2006).

values for the Gloves Task were 3.63 and 0.23. The three curves (from Jordan & Brannon, Gloves, and Cups) were clearly different from the curve obtained by Droit-Volet *et al.* The latter has a larger PSE (5.8) and lower sensitivity or, equivalently, a larger Weber ratio (0.32). The interpretation of the differences in the average curves is difficult because, in addition to the procedural differences already mentioned among the studies (e.g., simultaneous versus successive tasks, instructions not to count), we show next that in the present experiment the average functions do not represent well the data from all children.

Figure 2 shows the individual psychometric functions obtained in the Gloves Task. In each panel, the symbols show the data and the line shows the best-fitting logistic function. Based on the estimated slopes, the 15 psychometric functions seem to fall into two distinct groups, one in which the percentage of “many” choices increases gradually with the number of drumbeats (the top two rows of panels), and the other in which it increases abruptly, in a step-like manner (bottom row). For the former, group *Gradual*, the slopes ranged from 1.1 (S10) to 4.8 (S6) with a mean equal to 2.3, whereas for the latter, group *Step*, the slopes were all greater than 20. In fact, 4 children in group *Step* chose the “many” cup after all numerosities greater than 2 and 1 child (S8) chose it after all numerosities greater than 3. The logistic curve accounted for 88 to 100 percent of the variance in the data from group

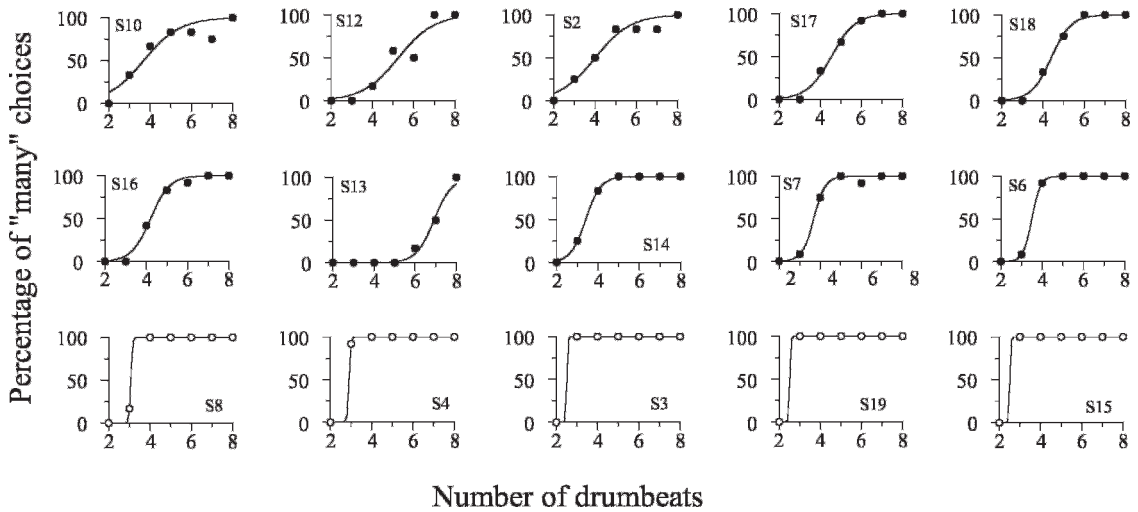


Fig. 2. Psychometric functions (symbols) and best-fitting logistic curve (line) for each child in the Gloves Task of Experiment 1. The functions in the bottom row (Step group) have much higher slopes than the functions in the top and middle rows (Gradual group).

Gradual (mean = 97%) and for 100 percent in the data from group Step.

The classification of the psychometric functions into two groups according to their slopes should be interpreted with caution for two reasons. First, dividing the groups is to some extent arbitrary. For example, participants S6 (slope 4.8) and, to a lesser extent, S7 (slope 3.5) could have been included in group Step. Second, not much weight should be attached to the specific values of the slopes for group Step because the slope of a step function at the step is infinite. What is important to note is the fact that the slope of the psychometric functions increased gradually for some children and abruptly for others.

The two types of functions differed not only in the slope but also in the PSE parameter. The mean PSE for group Gradual was 4.4 (95 percent confidence interval: 3.6 – 5.1), and the mean PSE for group Step was 2.7 (2.5 for 3 children, 2.9 and 3.0 for the other 2).

Figure 3 shows the individual psychometric functions for the Cups Task. As before, the functions seem to fall into two distinct groups according to their slopes: group Gradual (top three rows, with slopes ranging from 0.7 [S5] to 2.8 [S9] with an average of 1.5), and group Step (bottom row, with one slope equal to 7.3 [S17] and all others greater than 20). The mean PSE also differed substantially between the two groups. In group Gradual the mean

equaled 4.7 (95% confidence interval: 4.4 – 5.0) and in group Step it equaled 2.8 (3.5 for 3 children and 2.9 and 3.8 for the other 2).

Fourteen children completed both tasks. To compare their performances across tasks, Figure 4 shows the two psychometric functions in the same panel. For 4 children (see bottom row), the two functions were step-like and basically overlapped, which means that their performances in the two tasks were identical. For the remaining children, one or both functions increased gradually but with the exception of one child (S18) the functions did not overlap. For 6 children (top five panels plus the first panel in the middle row), the function for the Gloves Task was to the left of the function for the Cups Task, but for the remaining 3 children (S17, S12, and S13), the opposite was the case. With the exception of child S18, the curves in the two tasks differed in terms of slope, location, or both parameters.

In summary, in both tasks, some children (group Step) seemed to make a categorical discrimination, choosing “few” after two or three drumbeats and “many” after more than two or three drumbeats, whereas other children (group Gradual) made a noncategorical discrimination, increasing gradually the tendency to choose “many” with the number of drumbeats. The bisection points differed between the two groups, being smaller and

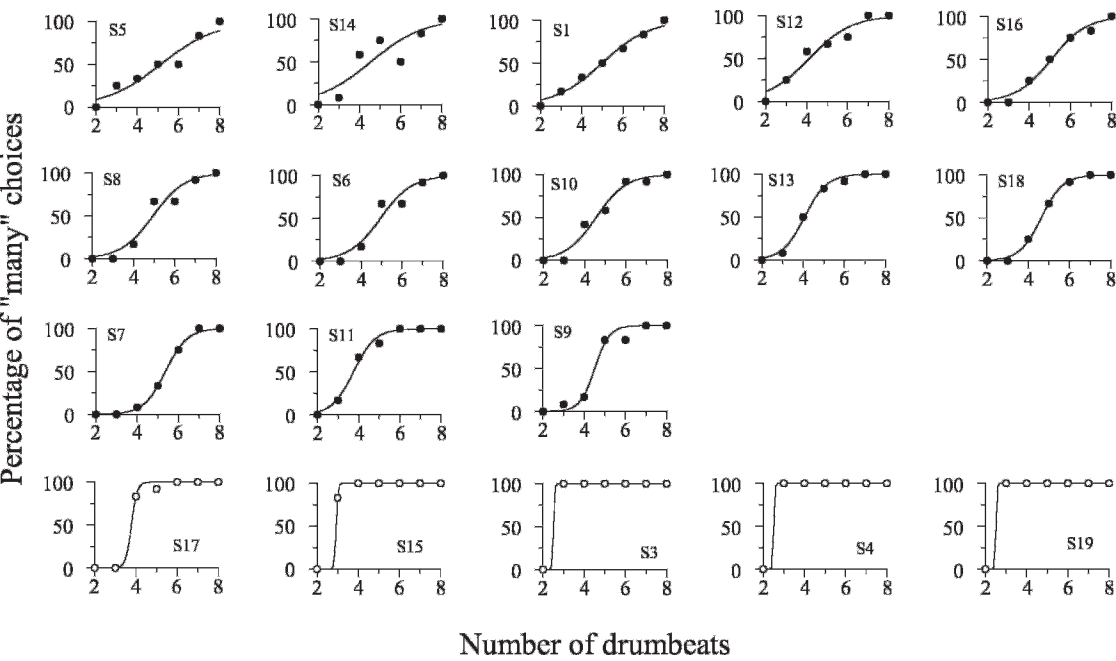


Fig. 3. Psychometric functions (symbols) and best-fitting logistic curve (line) for each child in the Cups Task of Experiment 1. The functions in the bottom row (Step group) have much higher slopes than the functions in the other rows (Gradual group).

clearly below 4 in group Step and larger and generally at or above 4 in group Gradual. All children counted correctly the number of coins earned during the experiment and

responded correctly to the experimenter's requests for three coins, a pile with five coins and another with eight coins. Moreover, spontaneous verbalizations during the tasks

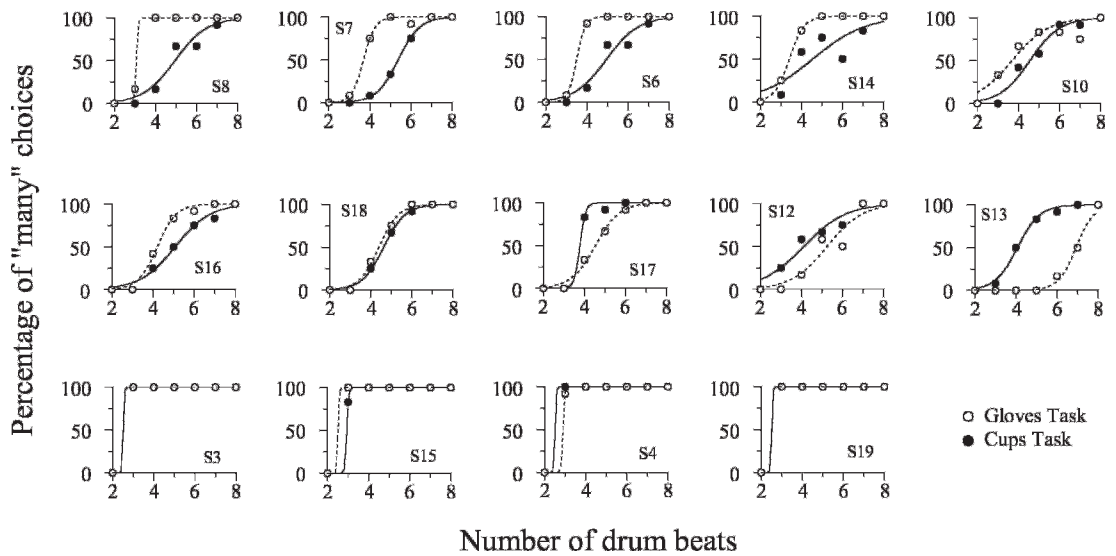


Fig. 4. Psychometric functions (symbols) and best-fitting logistic curve (line) for each child who completed both the Gloves (open circles) and Cups (filled circles) tasks of Experiment 1.

included statements such as the following: "The Red glove is when you beat four times and the Yellow is when you beat nine times"; "[the same child in the Cups Task] Green is when you beat two times and Blue five"; "Green is four times and Blue is eight"; "Red is low and Yellow is high"; "When it is a little it is the Red glove and when it is a lot it is the Yellow"; "Green is very little sound, and blue is when it [the sound] is big"; "Blue is many times and Green is few times"; "Blue is when you beat a lot and Green when [you beat] a little bit". These verbalizations are highly variable. Some mention specific numbers (often incorrectly!), others mention approximate quantifiers, either numeric ("many", "few") on nonnumeric ("a lot of sound", "big sound"). How these verbalizations might relate to choice performance is examined below.

EXPERIMENT 2

All children who completed the Cups Task in Experiment 1 learned a new discrimination in Experiment 2, to associate four drumbeats with one cup and 16 drumbeats with the other cup. The assignment of the two cups to the two numerosities preserved the order of Experiment 1 (e.g., if the blue cup was correct after two drumbeats in Experiment 1, then it was correct after four drumbeats in Experiment 2). During test trials, the experimenter introduced samples with intermediate numerosities. At issue was the superimposition of the two psychometric functions, one obtained in Experiment 1 with numerosities ranging from 2 to 8, and the other obtained in Experiment 2 with numerosities ranging from 4 to 16. To determine whether the prior learning of the '2 vs. 8' discrimination affected terminal performance in the '4 vs. 16' discrimination, a new group of children without any previous training was included in Experiment 2.

METHOD

Subjects

The sample consisted of 24 children, 11 girls and 13 boys, with a mean age of 5 years and 10 months (range: 5 years and 5 months to 6 years and 4 months). Eighteen of these children participated in the Cups Task of Experiment 1. The other 6 were experimen-

tally naïve and came from the same two preschools.

Materials and Procedure

The materials and the procedure were the same as in the *Cups Task* of Experiment 1, except that the children learned to discriminate between 4 and 16 drumbeats. The intermediate numerosities used during the generalization test were 6, 8, 10, 12 and 14.

RESULTS AND DISCUSSION

All children learned the basic discrimination and chose always the correct alternative during the last five trials with each sample. The left panel of Figure 5 shows the average data from the two groups of children, those with and those without previous training. The two functions were monotonically increasing with the PSE (8.4 and 8.6) close to the geometric mean of the two training numerosities (8). The functions essentially overlapped. A two-factor, between-within ANOVA comparing the performance of the two groups of children (with and without previous training as the between factor) and seven numerosities (4, 6, ..., 16 as the within factor) revealed a strong effect of numerosity, $F(6,132) = 82.8$, $p < .001$, but no effect of group or the interaction between the two factors ($F < 1$).

The right panel of Figure 5 shows the average data from the '4 vs. 16' discrimination (Experiment 2) compared with the average data from the '2 vs. 8' discrimination (Experiment 1). To compare the two data sets directly, all stimulus numerosities in the '4 vs. 16' task were divided by 2. The results show that the two average functions were monotonically increasing, had PSEs close to 4 (4.1 and 4.2), and overlapped. A two-factor, repeated measures ANOVA comparing performance on the two ranges and seven (scaled) numerosities revealed a strong effect of numerosity, $F(6,102) = 131.8$, $p < .001$, but no effect of range or interaction ($F < 1$).

We conclude that at the group level, the performance of children with previous training did not differ from the performance of children without previous training. In addition, mean performance in the '4 vs. 16' discrimination was a scale transform of mean performance in the '2 vs. 8' discrimination. This result reproduces in a successive task the

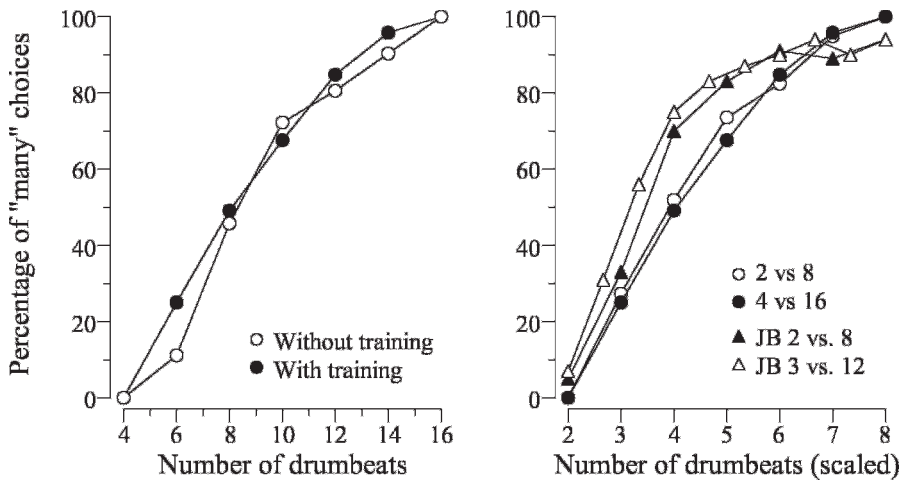


Fig. 5. Left panel: Average psychometric functions from Experiment 2 for the children with (filled circles) and without (open circles) previous training. Right panel: Average psychometric functions from Experiments 1 ('2 vs. 8') and 2 ('4 vs. 16') and from two conditions of Jordan & Brannon's (2006) study, abbreviated JB (2006), one condition with '2 vs. 8' and another condition with '3 vs. 12'.

result reported by Jordan and Brannon (2006) in a simultaneous task. However, we show next that the average psychometric functions did not represent well all individual functions.

Figure 6 shows the individual psychometric functions and the best-fitting logistic curves for the children with previous training. As in Experiment 1, the functions fall into two groups according to the estimated slopes. The panels in the bottom row correspond to group Step in which $\lambda > 9$. The panels in the other rows correspond to group Gradual in which $0.5 \leq \lambda \leq 2.5$ (average = 1.0). Again, the precise value separating the two groups is to some extent arbitrary but the existence of two distinct types of function is not. The group PSE averaged 8.4, a value close to the geometric mean of 4 and 16. If one considers only group Gradual, the average PSE was 8.7 and the 95 percent confidence interval ranged from 7.6 to 9.8. The 3 children from group Step had PSEs equal to 5, 6, and 10.

Figure 7 shows the data for the 6 children without previous training. The curves for the first 5 increased gradually with the number of drumbeats ($0.5 \leq \lambda \leq 1.5$, average $\lambda = 1.1$; average PSE = 9.5; 95% confidence interval = 6.2 – 12.7), whereas the curve for one child (S24) increased abruptly ($\lambda > 19$; PSE = 7.0).

The psychometric functions of the children with and without previous training were

similar. In both groups, there were two types of functions, one that increased gradually with stimulus number (the majority), and another that increased abruptly. Concerning the specific values of the parameters, neither the PSEs nor the slopes differed significantly between the two groups (PSEs: $t(22) = 0.6, p = .55$; slopes: $t(18) = 0.09, p = .93$; the analysis of the slopes was restricted to the gradually increasing functions—hence the difference in the degrees of freedom of the two t -tests).

To determine whether the performances of individual children in the '4 vs. 16' and '2 vs. 8' tasks were scale transforms of each other, Figure 8 plots the two psychometric functions and their best-fitting logistic curves in the same panel. The filled and open circles correspond to the 2 to 8 and 4 to 16 ranges, respectively, with the latter range rescaled by dividing all numerosities by 2. With respect to the parameters of the logistic, the rescaling of the 4 to 16 range amounts to halving μ and doubling λ . For 8 children (top two rows of panels), the two psychometric functions overlapped considerably or had similar slopes and locations. However, for the remaining 10 children (bottom two rows), the two functions had distinctly different slopes or locations. There was substantial variability within children to the point that some children (e.g., S4 and S19 in the bottom row) produced in one task a step

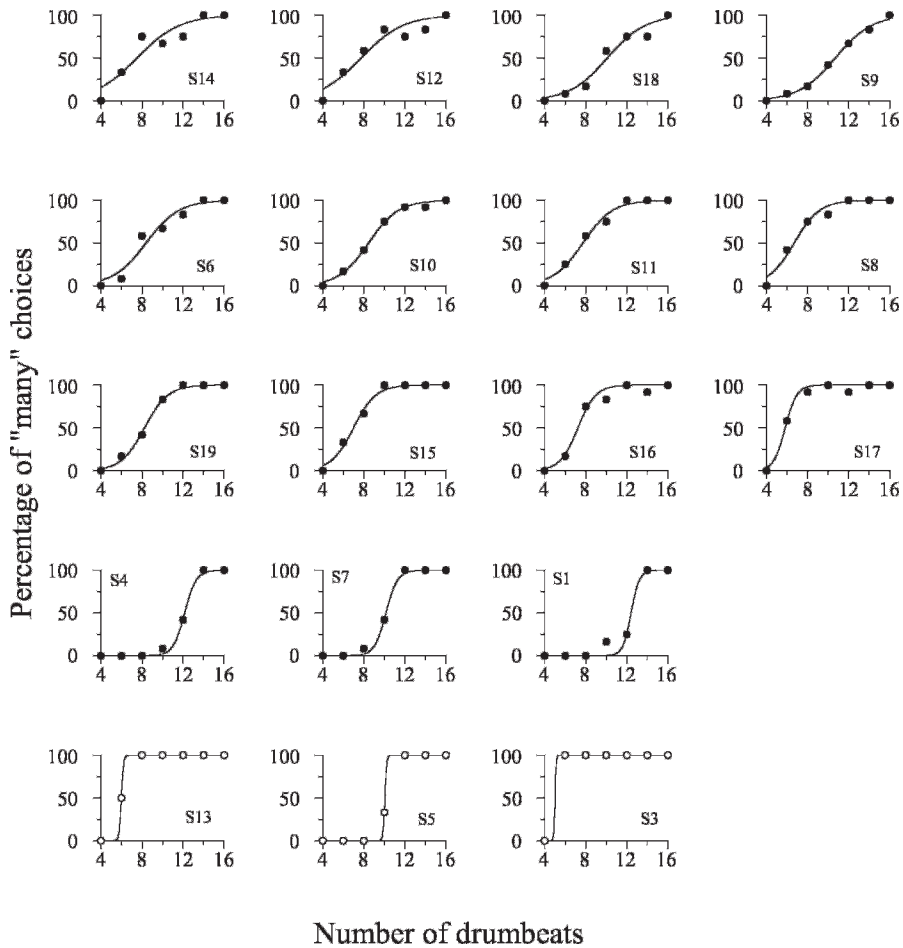


Fig. 6. Psychometric functions (symbols) and best-fitting logistic curve (line) for each child with previous training in Experiment 2. The functions in the bottom row (Step group) have much higher slopes than the functions in the other rows (Gradual group).

function (high slope) located at a small stimulus numerosity (low PSE) and in the other task a gradually increasing function (low slope) located at a large stimulus numerosity (high PSE).

To assess the statistical significance of the differences in the slopes and locations of the two functions, *t* tests for related samples were performed. Concerning the PSEs, the average in the '2 vs. 8' task equaled 4.16 and in the '4 vs. 16' task it equaled 4.21, a difference not statistically significant [$t(17) = 0.16, p = 0.87$]. Concerning the slopes (for the Gradual functions only), the averages 1.48 ('2 vs. 8') and 1.88 ('4 vs. 16') also did not differ significantly [$t(10) = .85, p = .41$].

Spontaneous verbal statements were similar in kind to those obtained during Experiment 1. Some mentioned specific numbers, although incorrectly (e.g., "Green is when it is 3 and Blue when it is 11"); some mentioned approximate quantifiers, either numeric (e.g., "I think it is Blue when you beat more and Green when you beat less") or nonnumeric (e.g., "Green is slowly; Blue is strong"; "The longest sound is Blue; the short is Green"). There also were statements that mixed specific numbers and approximate quantifiers ("e.g., "I know! When it is a lot it is Blue... [At the end of the experiment] A lot is Blue, Green is a little, 3 or 2"; "Blue is when it is many times, 9 or 10, and Green when it is few times, 5, 2, or

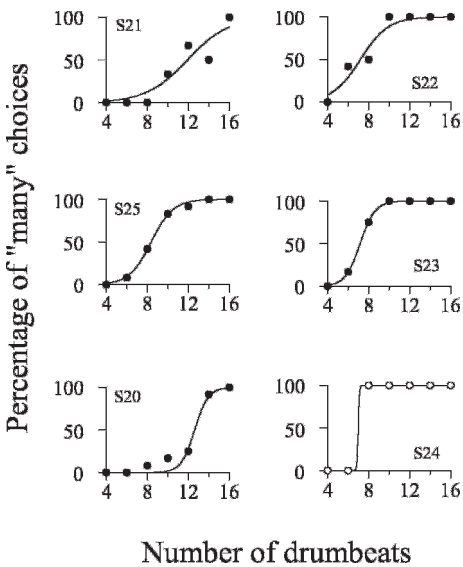


Fig. 7. Psychometric functions (symbols) and best-fitting logistic curve (line) for each child without previous training in Experiment 2. The function in the bottom right panel (Step group) has a much higher slope than the functions in the other panels (Gradual group).

3’’). No child named the correct numerosities of 4 and 16.

GENERAL DISCUSSION

We examined preschool children’s numerosity discrimination, presumably the lowest level of numerical competence. Children learned to discriminate 2 from 8 or 4 from 16 drumbeats without specific instructions regarding the numerical attributes of the sample or which behaviors to emit during the sample (e.g., counting). Next, to obtain psychometric functions, the children were exposed to stimulus generalization tests with intermediate numerosities. We analyzed three main issues. First, what are the properties of the psychometric function? Second, are psychometric functions obtained with numerosity pairs with the same ratio scale transforms of each other? Third, do group psychometric functions represent well the data from individual children? If not, then how do we describe and account for between- and within-

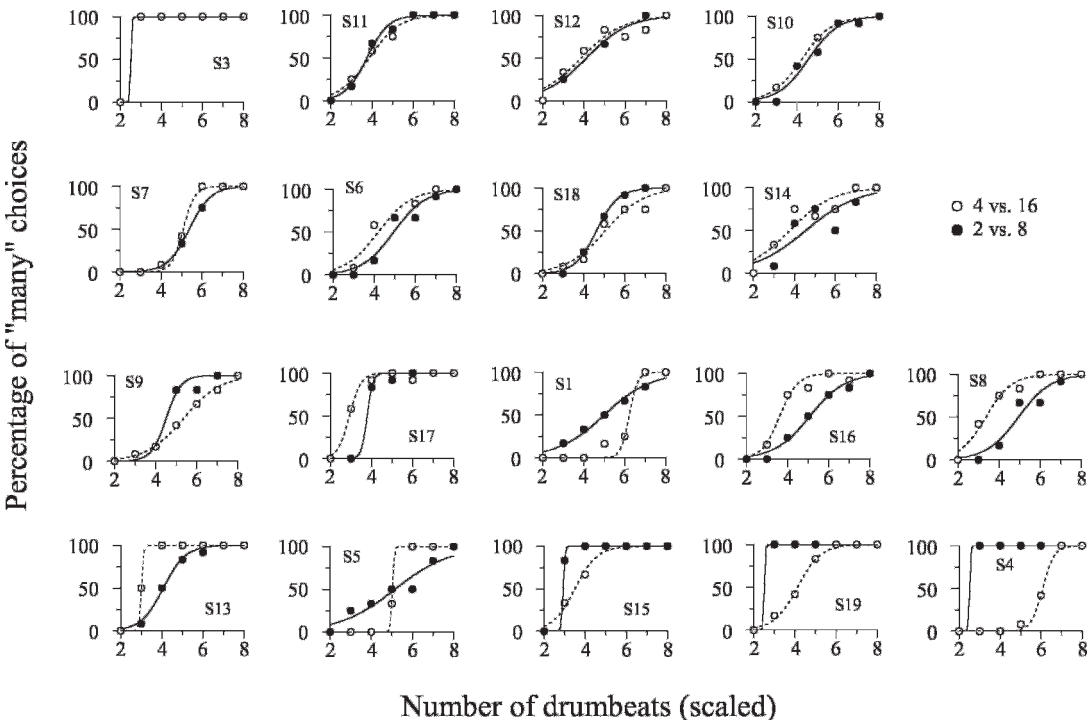


Fig. 8. Psychometric functions (symbols) and best-fitting logistic curve (line) for each child who performed the ‘2 vs. 8’ discrimination in Experiment 1 (filled circles) and the ‘4 vs. 16’ discrimination in Experiment 2 (open circles).

subjects variability? We address each of these issues next.

For each discrimination pair, the average of the psychometric functions was monotonically increasing with negative acceleration and had a PSE close to the geometric mean of the anchor numerosities. Similar PSEs were obtained by Jordan and Brannon (2006) using a simultaneous discrimination task with 6-year-old children and rhesus monkeys. In contrast, Droit-Volet et al. (2003) obtained PSEs close to the arithmetic mean using a successive discrimination task similar to ours with 5-year-old children (see Figure 1). Although we do not know how to account for the difference in the PSEs, we can rule out the hypothesis stated in the Introduction that successive tasks yield a PSE close to the arithmetic mean whereas simultaneous tasks yield a PSE close to the geometric mean, because our results with a successive task yielded an average PSE close to the geometric mean.

The results from Experiments 1 and 2 showed that the average psychometric functions for the '2 vs. 8' and '4 vs. 16' discriminations were scale transforms and therefore overlapped when plotted on a common scale (see Figure 5). The superimposition of the two mean functions reproduces in a successive task the superimposition results reported by Jordan and Brannon (2006) with a simultaneous task. The superimposition also reproduces the standard results obtained with animals (e.g., Fetterman, 1993; Fetterman, Dreyfus, & Stubbs, 1985; Gallistel, 1990; Meck & Church, 1983).

Our average curves are similar to the average curves obtained in other studies (Droit-Volet et al., 2003; Jordan & Brannon, 2006), but we also found that the average data do not represent well all individual data. In both experiments, a small but sizeable group of children produced step functions, which suggest a categorical discrimination between the two anchor numerosities, whereas the other children produced gradually increasing functions consistent with a noncategorical discrimination. The two types of functions were observed in both tasks of Experiment 1 (Gloves and Cups), in both groups of children in Experiment 2 (with and without previous training), and in both numerosity ranges within the same child (see Figure 8, bottom row). Whether these two types of functions

were present also in the two other studies with children (Droit-Volet et al.; Jordan & Brannon) is not known because individual data were not reported.

With respect to variability between children (task and the numerosity range held constant), the analysis suggested that the distribution of the PSEs differs in the two groups of functions. On the one hand, the 43 functions included in group Gradual had PSEs ranging from below the geometric mean to above the arithmetic mean. The averages of the PSEs were always between the geometric and arithmetic means and the 95 percent confidence intervals included both means in one case (Experiment 1, Gloves Task), neither the geometric mean nor the arithmetic mean in another case (Experiment 1, Cups Task), and only the geometric mean in yet another case (Experiment 2, including children with and without previous training). On the other hand, of the 14 functions classified in group Step (both experiments included), 13 had a PSE below the geometric mean of the anchor numerosities and only 1 (see S5 in Figure 6) had a PSE close to the arithmetic mean. It seems that categorical discriminations, expressed as step functions, yield smaller PSEs than noncategorical discriminations, expressed as gradually increasing functions. The latter in turn did not provide conclusive evidence concerning the (geometric mean or arithmetic mean) location of the PSE.

With respect to within-subjects variability, the analysis revealed that, for about half of the children, the psychometric function changed markedly with task and numerosity range. Thus, of the children who performed both the Cups Task and Gloves Task in Experiment 1, about half produced psychometric functions that varied in slope, PSE, or both (see Figure 4). Similarly, of the children that completed the '2 vs. 8' and '4 vs. 16' tasks, about half produced clearly non-overlapping functions; again, the (scaled) slope, the (scaled) PSE, or both varied with the numerosity pair. Interestingly, a few children produced a step function with one numerosity pair and a gradual function with the other.

The causes of the between- and within-subjects variability in the psychometric functions remain to be investigated. The general ability to count does not seem to be one of them because when the children were asked to

count the coins they earned and form piles with a specific number of coins at the end of the first task in Experiment 1, all of them performed correctly, which means that the ability to count does not seem to determine what children will do during the generalization test or, more specifically, the properties of their psychometric functions. However, differences in verbal behavior between and within children may account at least in part for the differences in the psychometric functions. The present study was not designed to examine the relations between numerosity discrimination and verbal behavior, let alone to disentangle which is cause and which is effect, but the following remarks may pave the way for such studies.

Humans may share with animals a basic, primitive sensitivity to stimulus numerosity, a sensitivity that, obviously, is not verbally mediated and that is expressed in the Gradual functions. This primitive sensitivity may be described at least in part by models such as Meck and Church's (1983) accumulator model developed on the basis of Gibbon's Scalar Expectancy Theory of timing (see Dehaene, 1997; Gallistel & Gibbon, 2002; Gibbon, 1977; Meck, Church, & Gibbon, 1985; Roberts, 1998). According to this model, animals have a neural module that consists of a pacemaker that generates pulses at a high rate, an accumulator that counts the pulses emitted during the to-be-counted sample, and one or more long-term memory stores that save the counts obtained at the end of each stimulus. In a bisection task, the model assumes the animal forms two memory stores, one containing the counts obtained at the end of the small numerosity sample (N_F) and the other the counts obtained at the end of the large sample (N_M). To decide which alternative to choose at the end of a sample, the animal compares the number that is in the accumulator when the sample ends (i.e., the number of pulses generated during the sample, N_T) against two samples, one extracted from the memory store for small numerosities, N_F , and the other from the memory store for large numerosities, N_M . If the ratio N_F/N_T is greater than the ratio N_T/N_M , then the number is "closer" to the sample extracted from the small numerosity store and the animal is more likely to choose the "few" key; otherwise, the animal is more likely to choose the "many" key. The model

predicts the S-shaped form of the psychometric function, the PSE at the geometric mean of the training numerosities, and the superimposition of the functions obtained with numerosity pairs in the same ratio when scaled appropriately.

During development children learn to verbally categorize their environment and, more specifically, tact the sample stimulus and count. As a consequence, the primitive sensitivity to numerosity described above may be replaced under some circumstances by behavior that is verbally mediated and that is expressed in the Step functions. Accumulator models would remain appropriate only when counting or verbal mediation of choice did not occur (e.g., with very large numerosities or short reaction times; Dehaene, 1997).

But while preschool children become more apt at counting and verbally categorizing stimuli, they may not do it equally well, or equally consistently, across tasks and numerosity ranges (e.g., children may have greater difficulties sustaining attention with larger numerosities or counting rapidly large sequences of drumbeats). These potential sources of variability may explain the occurrence of Gradual and Step functions in different children performing the same task, and in the same child performing different tasks (e.g., Cups and Gloves) or performing the same task but with different numerosity ranges (2 to 8 or 4 to 16).

The hypothesis that humans may share with animals a primitive sensitivity to numerosity, which later may be replaced by verbally mediated behavior, also helps to explain what otherwise could be a puzzling fact: The two numerosity tasks can be solved easily by choosing the "few" comparison stimulus following 2 (Experiment 1) or 4 (Experiment 2) drumbeats and the "many" comparison stimulus following all other samples. Given the dichotomous nature of the training, and the fact that reinforcement was available for all responses during testing, the children could have learned this simple discrimination and then maintained it on test trials. In retrospect, then, it may be surprising that only a minority of children produced Step functions. The fact that the majority of children did not act in this way and instead seem to take both numerosities into account—as pigeons and rats typically do—suggests that sharp, digital-like, cate-

gorical discriminations are not easier or more primitive than fuzzy, analogical-like, continuous discriminations (see Dehaene, 1997).

The foregoing hypothesis is consistent with the variety of the content of children's verbalizations. As mentioned above, some children used exact quantifiers (e.g., "Two is the Green cup"), whereas others made statements that used approximate, inexact quantifiers, either numeric ("Few is the Green cup") or nonnumeric ("The longest sound is Blue; the short is Green"). Some children mixed the two types of numeric quantifiers ("Two is Green, many is Blue") or even the numeric and nonnumeric quantities (e.g., "Green is few sounds and Blue is a lot, when it was big"). Even when the child verbalized two numbers, she did not report the correct numerosities! The verbalization of nonnumeric quantities and of inexact quantifiers, and the absence of precise control by number even when number is verbalized, all suggest that the sensitivity to numerosity is, at least initially, similar to sensitivity to other stimulus dimensions (e.g., amount—see Mix et al., 2002; time—see Meck & Church, 1983).

This last remark raises the issue of whether performance in the present experiment was determined by sample duration or sample numerosity. The present study did not attempt to disentangle the two sources of control, but three facts suggest that numerosity played the larger role. First, the average results as well as the results from the majority of the psychometric functions (the gradual functions) are consistent with the results obtained in numerosity discrimination studies with animals and children that controlled for nonnumeric attributes (e.g., Jordan & Brannon, 2006; Meck & Church, 1983; see also Roberts & Mitchell, 1994). Second, the study by Droit-Volet et al. (2003) showed that in preschool children, number is more likely to contaminate temporal judgments than time is to contaminate numeric judgments. This finding suggests that number is a more salient stimulus dimension than time. Third, many verbalizations referred explicitly to number or numeric attributes whereas only one referred explicitly to sample duration.

In conclusion, the present study showed that simple bisection procedures such as the Gloves and Cups tasks are useful for studying the numerical competence of young children. It

also showed that at a group level, numerosity discrimination in preschool children is similar to numerosity discrimination in animals. However, the study also revealed substantial variability between children in the same task and within children across tasks and numerosity ranges. One source of this variability may be the children's verbal behavior in general and the verbal quantifier they use to frame a decision rule in particular. Hence future research should examine directly the variables that influence children's decision rule. It may be the case that different tasks and contexts evoke different verbalizations and decision rules, and these in turn may lead to systematically different slopes and PSEs. If this hypothesis is correct, then the questions would no longer be whether the PSE is at the geometric mean or the arithmetic mean, or whether the psychometric functions overlap, but rather under which conditions the PSE takes this or that value, or under which conditions sensitivity to numerosity follows Weber's law.

REFERENCES

- Brannon, E. M., & Roitman, J. D. (2003). Nonverbal representations of time and number in animals and human infants. In W. H. Meck (Ed.), *Functional and neural mechanisms of interval timing* (pp. 143–182). New York: CRC Press.
- Davis, H., & Perusse, R. (1988). Numerical competence in animals: Definitional issues, current evidence, and a new research agenda. *Behavioral and Brain Sciences*, 11, 561–615.
- Dehaene, S. (1997). *The number sense*. New York: Oxford University Press.
- Droit-Volet, S., Clément, A., & Fayol, M. (2003). Time and number discrimination in a bisection task with a sequence of stimuli: A developmental approach. *Journal of Experimental Child Psychology*, 84, 63–76.
- Emmerton, J. (2001). Birds' judgments of number and quantity. In R. G. Cook (Ed.), *Avian visual cognition*. Retrieved July 7, 2007, from Tufts University, Department of Psychology Web Site: <http://www.pigeon.psy.tufts.edu/avc/emmerton/>
- Fetterman, J. G. (1993). Numerosity discrimination: Both time and number matter. *Journal of Experimental Psychology: Animal Behavior Processes*, 19, 149–164.
- Fetterman, J. G., Dreyfus, L. R., & Stubbs, D. A. (1985). Scaling of response-based events. *Journal of Experimental Psychology: Animal Behavior Processes*, 11, 388–404.
- Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.
- Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, 44, 43–74.
- Gallistel, C. R., & Gibbon, J. (2002). *The symbolic foundations of conditioned behavior*. Mahwah, NJ: Erlbaum.

- Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.
- Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, 84, 279–325.
- Gibbon, J. (1981). On the form and location of the psychometric bisection function for time. *Journal of Mathematical Psychology*, 24, 58–87.
- Honig, W. K., & Matheson, W. R. (1995). Discrimination of relative numerosity and stimulus mixture by pigeons with comparable tasks. *Journal of Experimental Psychology: Animal Behavior Processes*, 21, 348–363.
- Honig, W., & Stewart, K. (1989). Discrimination of relative numerosity in pigeons. *Animal Learning & Behavior*, 17, 134–146.
- Jordan, K. E., & Brannon, E. M. (2006). A common representational system governed by Weber's law: Nonverbal numerical similarity judgments in 6-year-olds and rhesus macaques. *Journal of Experimental Child Psychology*, 95, 215–229.
- Keen, R., & Machado, A. (1999). How pigeons discriminate the relative frequency of events. *Journal of the Experimental Analysis of Behavior*, 72, 151–175.
- Machado, A., & Keen, R. (2002). Relative numerosity discrimination in the pigeon: Further tests of the linear-exponential-ratio model. *Behavioural Processes*, 57, 131–148.
- Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, 9, 320–334.
- Meck, W. H., Church, R. M., & Gibbon, J. (1985). Temporal integration in duration and number discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, 11, 591–597.
- Mix, K. S., Huttenlocher, J., & Levine, S. C. (2002). *Quantitative development in infancy and early childhood*. New York: Oxford University Press.
- Piaget, J. (1952). *The child's conception of number*. London, England: Routledge & Kegan Paul.
- Rilling, M. (1967). Number of responses as a stimulus in fixed interval and fixed ratio schedules. *Journal of Comparative and Physiological Psychology*, 63, 60–65.
- Rilling, M. (1993). Invisible counting animals: A history of contributions from comparative psychology, ethology, and learning theory. In S. T. Boysen, & E. J. Capaldi (Eds.), *The development of numerical competence: Animal and human models* (pp. 3–37). Hillsdale, NJ: Erlbaum.
- Rilling, M., & McDiarmid, C. (1965, April 23). Signal detection in fixed-ratio schedules. *Science*, 148, 526–527.
- Roberts, W. A. (1998). *Principles of animal cognition*. London: McGraw-Hill.
- Roberts, W. A., & Mitchell, S. (1994). Can a pigeon simultaneously process temporal and numerical information? *Journal of Experimental Psychology: Animal Behavior Processes*, 20, 66–78.
- Shettleworth, S. J. (1998). *Cognition, evolution and behavior*. New York: Oxford University Press.
- Sidman, M. (1960). *Tactics of scientific research*. New York: Basic Books.

Received: March 5, 2007

Final acceptance: July 12, 2007